Jurnal Ekonomi dan Bisnis Indonesia
Vol. 16, No. 4, 2001, 358 - 371

ROBUST METHOD: AN APPLICATION TO DETERMINANTS OF RESEARCH AND DEVELOPMENT EXPENDITURES TESTING MODEL

Suwardjono
Universitas Gadjah Mada

ABSTRAK

Penelitian-penelitian akuntansi yang menggunakan pengujian statistis parametrik biasanya berasumsi bahwa populasi berdistribusi normal. Hasil pengujian parametrik yang didasarkan pada normalitas data mulai banyak dipertanyakan validitasnya karena kenyataan bahwa data bisnis dan akuntansi pada umumnya tidak memenuhi asumsi kenormalan. Yang menjadi pertanyaan adalah apakah mean dan variansi sampel masih tetap andal dijadikan estimator parameter populasi normal sementara pendekatan nonparametrik tidak memungkinkan untuk mengestimasi parameter.

Makalah ini mengevaluasi kemampuan metoda tegar sebagai alternatif alat statistis yang menuntut kenormalan data. Evaluasi dilakukan dalam rangka pengujian empiris faktor-faktor yang menentukan tingkat pengeluaran untuk riset dan pengembangan. Hasil empiris menunjukkan bahwa metoda tegar menghasilkan estimator parameter yang lebih konsisten dengan teori dan sekaligus mengatasi kelemahan yang ditimbulkan oleh model parametrik. Makalah ini merupakan seri makalah metoda statistis dari makalah sebelumnya (Suwardjono, 2001).

Kata kunci: *riset dan pengembangan, metoda tegar, kuadrat terkecil biasa, median terkecil regresi kuadrat, parametrik, nonparametrik, asumsi normalitas.*

INTRODUCTION

Research and development (R&D) have been the subject of many studies ranging from those that investigate the differences in spending behavior between certain industries to those that investigate market reactions to changes in R&D accounting disclosure requirements. Research and development play a very important role in the economy as indicated by the fact that the government sponsors many research projects to promote growth in particular industries. Some studies indicate that R&D activity average about 2.5 percent of sales and 50 percent of net income across the economy. At the level of the firm, industry and economy, the contribution of R&D to economic growth/productivity is positive and significant (Dukes, Dyckman and Elliott, 1980). A company has to report the R&D expenditures and activity in a certain way according to some permissible methods and these rules certainly affect the company's R&D spending strategy. Several studies addressed the economic consequence of reporting rules for R&D expenditures (Horwitz and Kolodny, 1981; Elliott *et al.*, 1984; and Shehata, 1991).

Grabowski (1968) examined the determinants of research expenditures (in terms of research intensity) in drugs, chemicals and petroleum refining industries. It was concluded that the interim differences in technology, product diversification, and availability of funds were all significant in explaining research intensity. Madden, McCullers and VanDaniker (1972) conducted a survey to determine whether R&D expenditures are sufficiently material to warrant greater disclosure. The results indicated that the level of R&D expenditures in responding firms was material in relation to net income.

Several factors are frequently associated with the behavior of managers in spending decisions including R&D decisions and strategies. These factors are usually firm-specific (e.g. size) or industry factor. Some variables or factors that are frequently cited in the studies of R&D spending determinants are size, availability of fund, riskiness (diversification), aggressiveness, inventive activity, growth in productivity, an increase in knowledge as evidenced by an increase in patents and available accounting methods. However, the firm's behavior with respect to the R&D expenditure are still not well defined. Moreover, most studies use parametric models (ordinary least square models).

This paper examines several factors that may have an effect on the R&D expenditures in conjunction with testing the appropriateness of robust method in lieu of the common parametric models. Instead of focusing on certain industries, this paper performs the analysis on the manufacturing industry in general. Complete description of robust method and guideline for choosing appropriate methods are given in the Appendix.

R&D DECISIONS AND HYPOTHESES

Several different models are used in many studies to test the effects of some variables on a variable of interest. In testing the validity of robust method, this paper uses a regression model to test if the selected independent variables representing the firms-characteristics are individually or as a whole associated with the response variable.

The response variable is the research and development expenditures of firms for the year under the study (1992). This variable is measure in actual dollar value spent by each firm for the R&D activities during 1992. The year 1992 is chosen because of data availability and the time when most firms fully applied the Statement of Financial Accounting Standard (FAS) No. 2. Financial Accounting Standard Board (FASB) defines *research* as a planned search or critical investigation aimed at discovery of new knowledge to develop new product or process or to improve existing product or process. *Development* is defined as the translation of research findings or other knowledge into a plan or design for a new product or process or for a significant improvement to an existing product or process. When absolute or nominal values are used, heteroscedasticity is repeatedly present and the scale effect tend to dominate the regression equation. Some studies avoid this problem by deflating the response and explanatory variables with size factor, for example, sale or asset (see e.g. Grabowski, 1968 and Dukes, Dyckman and Elliott, 1980). An alternative approach is to transform the dependent variable whenever possible under the model specifications. The procedure adopted in this paper is to estimate the regression model using least median of squares regression (LMS) method, as one kind of robust method, developed by Rousseeuw (1984). Instead of deflating by size variables, the size variables are treated as explanatory variables.

The first determinant of the R&D expenditures to be considered here is the size of a company. The theory behind this choice is that larger firms are financially better equipped to undertake large-scale R&D projects than are smaller firms. The results of empirical studies on the size effect are mixed. Grabowski (1968) shows that R&D is proportionately related to firm size. To evaluate the validity of this finding, sales and assets are used as proxies for size and included as explanatory variables in this paper. Sales is not necessarily correlated with asset because firms are different with respect to turnover or efficiency of asset utilization. Therefore, both variables are included in the model. It is hypothesized that he larger the firms, the higher the R&D expenditures.

The second explanatory variable is the availability of fund to finance the R&D activities. R&D may be financed from external sources. However, external financing may jeopardize the financial position of the firm and may lead to breaching some debt contracts (covenants). Managers in general are value-maximizer, therefore they tend to the avoid the actions that decrease the value of the firm. In other words, the ability to finance R&D with external funds is limited due to the risky nature of R&D projects. It can be said then that firms rely on internal fund for R&D ventures and the availability of internally generated funds becomes a critical factor. Cash flows generated by operation is a measure of fund availability. It is hypothesized that the higher the cash flow generated internally, the greater the firm's ability to invest in risky R&D projects thus the higher the R&D expenditures (Shehata, 1991).

Firms also commit to improve and replace facilities. Capital expenditure decisions are assumed to be made in conjunction with the R&D expenditure decisions. As far as fund availability is concerned, the R&D and capital expenditures may be complementary or competing. The results of previous studies are mixed. Shehata (1991) points out that no directional relationship can be posited between R&D and capital investments. However, because the fund for both activities is usually limited, this paper assumes that R&D competes with capital investment in the use of fund. Therefore, it is hypothesized that the higher the capital expenditure, the lower the R&D expenditures.

The capital structure and financial position of the company may affect the R&D decisions. Firms that are in the lower margin with respect to the financial riskiness tend to avoid actions that worsen the financial position. In the study of the choice of accounting method in the oil and gas industry, Malmquist (1990) states that riskier group of companies tend to have exaggerated variance in debt-to-equity ratio and therefore tend to choose method that stabilizes income. Debt-equity ratio is an important variable influencing the choice of method. With similar reasoning, it can be said that the higher the financial risk (the lower the debt-equity ratio) the higher the R&D expenditures. Another measure of financial riskiness is product diversification. Highly diversified firms are stronger to withstand the unfavorable outcomes of certain R&D projects. Therefore it is hypothesized that the higher the degree of diversification, the lower the risk inherent in R&D investments and higher R&D expenditure is also expected. One measure of diversification is the number of four-digit SIC industries in which the firm operates. Due to the data availability, this variable is not considered in this paper.

The Testing Model

The determinants of R&D expenditures can be examined empirically by the application of linear regression model. Regression analysis is widely used in studies with the objective of examining the determinants of some accounting variables. For example, Francis and Reiter (1987) apply this method to investigate the determinants of corporate pension funding strategy and Shehata (1991) uses two-stage linear regression to evaluate the determinants of R&D expenditures. To test the hypotheses in this paper, a linear regression model of the following form is to be estimated:

$$RD_i = \beta_0 + \beta_1 SALES_i + \beta_2 CAPEX_i + \beta_3 CFLOW_i + \beta_4 ASSET_i + \beta_5 DER_i + \epsilon_i$$

where the variables are defined and measured as described in Table 1.

Preliminary analysis of the data and residuals by ordinary least square (OLS) estimation using all available observations indicates that the distribution of the data and residuals do not comply with the OLS assumptions. In particular, OLS estimation suffers from nonnormality and nonconstant variance problems. Therefore, least median of squares regression (LMS) is used as the last attempt to deal with the problems. In fact, alternative procedures are also appropriate, for example robust regression or M-estimators (see details in Booth, 1985). The reason to use LMS is simplicity and ease of application. In the first run, the model is estimated by OLS and LMS for all selected sample firms and the results are compared. The firms with nonzero weight in the LMS are then used to estimate the model by OLS to obtain outputs for residual analysis.

SAMPLE SELECTION AND DATA

The sample firms are selected from the COMPUSTAT data tapes. Firm data for 1992 are extracted. Again, the year 1992 is chosen because of data availability and the time when most firms applied fully the Statement of Financial Accounting Standard (FAS) No. 2. Initially, all firms in the data tapes that have R&D and other variables values greater than zero are extracted. This results in 564 available firms representing all industries. For the purpose of this paper, only manufacturing firms will be examined. Out of the available data, 225 manufacturing firms are selected at random based on SIC codes while eliminating nonmanufacturing firms. The reason for reducing the number of observations is the concern over dominant number of firms in certain industries and the limitation of LMS program to handle data (maximum of 300 observations). Table 1 describes the notation, expected sign and measurement of the variables for the selected sample firms.

EMPIRICAL RESULTS

Table 2 summarizes the results of the OLS estimation using SAS program for 225 sample firms. The OLS estimation using PROGRESS program produces the same results. The reason to use SAS program is to obtain summary analyses (scatter plot, normal probability plot, and other diagnostics) that are not provided by PROGRESS.

A large portion (84%) of the variation in R&D expenditures is explained by the model as shown by R-squared value. High value of F statistic ($F=229.7$ with $p=0.0001$) suggests that the model fits the data and the explanatory variables as a whole are important in explaining the variation of R&D expenditures. The table also indicates that sales, capital expenditure and cash flows are statistically significant at 0.05 level with the signs consistent with predicted signs except for capital expenditure. This means that R&D expenditures are complementary to the capital expenditures instead of competing. Although the signs are as predicted, asset and debt-equity ratio are not statistically significant. The OLS estimates, however, suffer from some violations of OLS assumptions so that the results may not be reliable. Univariate analysis of the data shows that the data for each variable is not normally distributed. In particular, the residual analysis indicates that the disturbances are not normally distributed. However, in large samples the normality assumption is not critical because the sampling distribution of the estimators of the regression coefficients will still be approximately normal (Dielman, 1991). Since we have large enough sample size in this paper, this is not a serious violation to affect the results.

Table 1 Definition of Variables and Their Measurements

Variable	Sign	Definition
Dependent:		
RD (research and development expenditures)		Total research and development expenditures for 1992 as defined in the COMPUSTAT manu als. The values of this variable are the figures as reported on the COMPUSTAT tape.
Explanatory:		
ASSET (firm size)	+	Total tangible assets as reported on the COM PUSTAT tapes.
SALES (firm size)	+	Net sales dollar as reported on the COM PUSTAT tapes.
CFLOW (cash flows as a fund availability measure)	+	Cash flows generated in 1992 and measured as the total of income before extraordinary items and depreciation and amortization. Data are taken from COMPUSTAT tapes.
CAPEX (capital expenditures as a measure of expenditure decision)	-	Total amount of capital expenditures incurred by the firm as reported on COMPUSTAT tapes.
DER (debt-equity ratio as a measure of riskiness)	-	Total long-term debt divided by the book value of equity. Data are taken from COMPUSTAT tapes.

Table 2 Estimates of the Model Using OLS for 225 Selected Firms

Variable	Coefficient	Standard Error	t-ratio	P-value	Variance Inflation
Intercept	-21.251252	10.864949	-1.956	0.0517	0.00000
SALES	0.019064	0.005650	3.374	0.0009	10.34420
CAPEX	0.336519	0.070170	4.796	0.0001	8.22576
CFLOW	0.175893	0.037657	4.671	0.0001	5.41685
ASSET	0.000167	0.002947	0.057	0.9550	4.92578
DER	-11.931453	9.758053	-1.223	0.2227	1.04169
Adjusted R2 = 0.8362					
F=229.70 (p>0.0001)					

The more serious violation of OLS estimation is constant variance assumption. Each residual plot of residuals against each of the explanatory variable shows invariably V-mass pattern or in the case of DER, a diamond pattern. The V-mass pattern also markedly appears in the residuals plot against predicted values (see Figure 1 Panel A). Variance inflation indexes indicate that SALES variable contains some influential outliers (variance inflation factor = 10.34). All these residual analyses suggest that the OLS estimates suffer from severe heteroscedasticity. Because of this problem, hypothesis tests about the population parameters based on the OLS estimates may provide misleading results.

Table 3 Estimates of the Model Using LMS for 225 Selected Firms

Variable	Coefficient	Standard Error	t-ratio	P-value	Variance Inflation
Intercept/Constant	1.53342	0.33158	4.62459	0.00001	0.00000
SALES	0.01394	0.00049	28.53509	0.00000	8.49163
CAPEX	-0.01962	0.00739	-2.65489	0.00890	3.88750
CFLOW	0.04893	0.00583	8.38969	0.00000	7.19885
ASSET	0.00130	0.00028	4.66452	0.00001	5.67530
DER	-1.30270	0.26854	-1.12719	0.26169	1.04016
Adjusted R2 = 0.9886					
F=2398.4 (p>0.0001)					

Several attempts were made to alleviate the nonconstant residual variance. Transformations of response variable (RD) by taking natural logarithm, square, inverse or square root did not help much. V-mass pattern persists in the residual plot and in some cases systematic pattern appears instead. Transformed models using square root of ASSET and SALES as deflators are also estimated but the results did not significantly stabilize the residuals. Since some efforts to fix violations of the OLS estimations do not provide satisfactory results, as the last attempt, the model is estimated by least median of squares regression (LMS). This method simply diagnoses outliers and puts weights of zero on detected outliers and recalculates the estimates so that the residuals are stabilized. Table 3 presents the results of this method. After diagnosing outliers, this method runs reweighted least square with 139 non-zero weight points.

The LMS estimation results in higher R^2 than OLS method (0.9890 compared to 0.8399). The increase implies that the homoscedasticity assumption is very important. The LMS estimation has reduced the impact of outliers and hence provides more powerful statistics than the OLS estimation. F statistic also increases considerably suggesting that the model fits better the remaining data. As we can see from Table 3, SALES, CAPEX, CFLOW, and ASSET are all statistically significant at the 0.05 level with the signs as predicted. With OLS method, only SALES, CAPEX, and CFLOW are significant with inconsistent sign for CAPEX. The DER is not statistically significant in any of the methods. One explanation for the insignificance of the DER is that debt-equity ratio might not capture the riskiness of the firm or the R&D project portfolio. As has been mentioned before, level of diversification may reflect the riskiness of the firm conducting research projects and thus a more appropriate surrogate.

Tempel Gambar 1 di sini

Figure 1 Plots of Residual Against Predicted Values

In order to evaluate the validity of the LMS results, it is desirable to examine the characteristics of its residuals. Since PROGRESS program does not provide residual plots other than standard plot, the 139 non-zero point observations retained by the LMS method are used to estimate the model by OLS using SAS program. This estimation is equivalent to the LMS and is carried out to obtain the necessary diagnostics. Moreover, the standard plot provided by LMS is visually very difficult to analyze because many observations are hidden (see Figure 2).

The following analyses are based on the results of estimation for 139 sample firms using SAS program. The plot of residuals against predicted values indicates no apparent systematic patterns or V-mass even though Wilk-statistic shows that the distribution is not normal. Even though a large portion of the observations lie in the lower (left) section of the predicted value axis, the residuals appear to scatter randomly and evenly across the predicted values (Figure 1 Panel B).

Figure 2 Plots of Residual Against Predicted Values from PROGRESS Printout.

The plots of residuals against each of the explanatory variables also show similar patterns except for DER variable which still show a diamond shape. This unique pattern may be ignored since the DER variable is not statistically significant. From the variance influence factors, we find that no variable has an index greater than 10. It can be said that no overall impact of outliers is present. Therefore it is unlikely that the LMS estimates lead to misleading interpretations. Relying on the test statistics provided by LMS, we cannot reject all the null hypotheses of no impact except for riskiness factor and we conclude that firm size, capital expenditures, and fund availability are important determinants of R&D expenditures.

SUMMARY AND CONCLUSIONS

Because of the nature of business and economic data which generally violate the assumptions of OLS estimation model, hypothesis tests based on the OLS estimates may provide misleading results especially if the residuals are heteroscedatic. LMS is an estimation method to minimize the influence of outliers. This paper provides evidence that LMS estimation can be useful and more powerful when all possible methods under OLS to stabilize residual variance fail to provide satisfactory results.

The primary drawback of LMS estimation developed by Rousseeuw (1984) is that suspected outliers are given zero weight. This is the same thing as eliminating observations. Moreover, the LMS model in this paper has eliminated too may observations. Out of 225 observations, only 139 (62%) points were considered nonoutliers. As Booth (1985) points out, there are problems with the elimination. First, an extreme point may provide us with useful information of some sort or another. Second, we cannot be sure that a particular point is an outlier just because it deviates more than two standard deviations from the mean. One possible explanation for too many deletions is that there is a latent discriminating factor that is not taken into account in the model.

Subject to limitations of the LMS estimation, the main conclusion of the analysis is that sales, asset, capital expenditure, and cash flow are all significant in explaining R&D expenditures. These variables represent the size of the firm, capital expenditure strategy and availability of fund. In general, these results are consistent with those of previous studies. In particular, the results in this paper support the claim that R&D expenditure is competing with the capital expenditures in the use of available fund. When a company has to make a major capital expenditure, some R&D activity may have to be reduced. The results also support the hypothesis that size does matter even though several previous studies provide mixed result. One possible reason for the insignificance of size variable is that R&D activities of most companies are long-term definite program so that expenditures are independent of level of size-related factors in a particular period.

APPENDIX

ON ROBUST METHOD

The least square estimators and their generalization have been dominating for a long time. Andrews (1974) points out that the linear regression and other normality-assumption-based (e.g. ANOVA) procedures are the most frequently used procedures at the University of Toronto. More than a half of the uses of the statistical package (BMDP) are linear regression type of analysis and almost every discipline is making use of the procedures. It seems that least square methods have been satisfactorily serving the needs of users (academicians and applied statisticians). The least squares (LS) estimation is strictly based on the assumption that the measurement errors should be normally distributed. Huber (1972) calls this assumption a dogma of normality and states that the use of arithmetic mean had become almost sacred over the years. He remarks that the normal distribution was introduced by Gauss to suit the sample mean. With the normal distribution, the mean is often said to be the linear unbiased estimator of the expected value of the underlying population. Huber further argues that the dogma of normality is indeed still widespread because users misunderstand the Gauss-Markov theorem and central limit theorem (CLT). The LS estimation was developed with the idea that almost all of the statistical variability is due to the measurement errors or other extraneous factors (external variability). As Hogg (1979) noted, the underlying assumption of LS is that outliers arising from other than normal distribution are simply considered as bad data points.

In the late 1950's, the parametric results based on normality assumption began to be questioned (see e.g. Rey, 1978 and Staudte, Jr., 1980). The question that is often raised in this respect is whether the sample mean and the variances are still reliable estimators of normal parameters when the data sets do not strictly satisfy the assumption (which are the most common cases in real life data). The weakness of the LS estimation is that instead of looking outliers as inherent statistical variability of the data, it treats them as measurement errors and nuisance and consequently they should be eliminated. This means that the main interest in estimation diverts from that of finding the true value to that of finding combination of observations which on the average lies nearest to the true values. The empirical distribution of the sample may suggest better estimates than those provided by the classical least square methods. Huber (1972) specifically suggests that we look at actual error distribution and examine whether the data are compatible with a normal distribution and, if not, to develop a different theory of estimation. Instead of imposing linearity, normality, and unbiasedness, alternative robust methods of estimation should be developed. This is the reason for the emergence of robust techniques of estimation. Robust statistics are techniques that are insensitive to small deviation from classical assumption (especially normality) and yet powerful to specific factors under the test.

CLASSICAL ROBUST STATISTICS (NONPARAMETRIC ANALYSIS)

The development of nonparametric methods is basically a response to the problems with the classical normal parametric approach. Since the underlying distributions of population do not always (in fact very rarely do) meet the assumption of parametric tests, inferential procedures whose validity does not depend on the rigid assumption of classical models are needed. Nonparametric statistical procedures are the answers to the needs. As the name implies, these procedures are not concerned with population parameters but with other characteristics such as goodness of fit and tests of randomness. Since the validity of procedures does not depend on the functional from of the sampled population, nonparametric procedures are sometimes called distribution-free procedures. This distribution-free nature is cited as the main advantage of nonparametric methods over the classical parametric methods because the chance that they are improperly used is small (see e.g. Daniel, 1990). The procedures bring some relief for testing problems.

However, in some situations, nonparametric procedures result in less powerful tests than their parametric equivalents since nonparametric procedures utilize less information from the sample data. Actually, nonparametric approach does not solve the problem of robustness that classical least square method suffers from but rather they are different methods for different purposes. Unlike robust estimators, most of the techniques in nonparametric analysis are for hypothesis testing but not for the purpose of location estimation. Moreover, there are no strong general underlying principles in nonparametric procedures so that they have general and wide applications. In fact, as Huber (1972) maintains, the notions of nonparametric and distribution free have a very little thing to do with robustness in a real sense. We can say that sample mean and the sample median are nonparametric estimates of the true mean and median but we do not know with certainty which functional probability of the distribution we want to estimate. In summary, as far as the purposes (estimation, robustness, and power) are concerned, nonparametric procedures are not the alternatives of classical statistics but rather a different family of statistics with the aim of avoiding the problems of inherent statistical variability of the data (distribution) rather than dealing with them. By nature, nonparametric analyses are not concerned with the detection of outliers and probability distribution of the data.

MODERN ROBUST STATISTICS

Many test procedures involve probability and therefore depend for their results on the assumptions concerning the variation of the population. As discussed previously, the Gaussian or normality assumption is the most widely adopted assumption about the generating mechanism of the data. Parametric models work well when all the underlying assumptions are fully met. If inference are slightly affected by departure from those assumption (if the probability value of statistics is stable) the tests on the inferences said to be robust. Quoting the Kendal and Buckland dictionary, Staudte (1980) defines "statistical procedures are robust if they are not very sensitive to departure from the assumption on which they depend." Booth (1986) defines an unbiased estimators of population parameter is robust if a large change in one sample point produces only a small change in the estimate. Mallows (1979) defines robustness in terms of three attributes: resistance, smoothness, and breadth. *Resistance* refers to the properties of being insensitive to the presence of a moderate number of bad values in the data. *Smoothness* refers to improvement of the concept through gradual response to small errors and changes in the model and *breadth* refers to the applicability of methods in a wide variety of situations.

Unlike nonparametric analysis, the robust procedures are more concerned with estimation method as alternatives to the classical model. Many robust procedures have been introduced in the literature as a response to the inadequacy of classical parametric approach. Maximum likelihood estimators (M-estimators), linear combinations of order statistics (L-estimators), estimates derived from rank tests (R-estimators) and adaptive estimators are the major families of robust procedures (see discussions in Huber, 1972 and Hogg, 1979.).

Compared to the classical least square procedures, the main advantage of robust methods is their stability against the presence of outlying values whatever the source of the errors. Huber (1972) recommend the use of robust procedures over the parametric procedures for the following theoretical reasons:

- One never has a very accurate knowledge of the true underlying distribution.
- The performance of some of the classical tests or estimates is very unstable under small change of the underlying distribution.
- Some alternative tests or estimates lose very little efficiency for an exactly normal distribution model but show a much better and more stable performance under deviation from that model.

ADVANTAGES AND DISADVANTAGES

Advantages of Nonparametric Statistics:

- Nonparametric statistics is not based on classical assumptions about errors but on minimum assumption and therefore they have a little chance of being improperly used.
- Formulas are not mathematically involved and can be easily applied especially far small samples.
- The concepts and procedures are usually easy to understand for those with minimum background. However, it does not mean that the procedures are easy to develop.
- The procedures can be applied to analyze count or rank data.

Disadvantages of Nonparametric Statistics:

- Because of minimum assumptions, the procedures are often misapplied for analyses which require more stringent assumptions so that the validity of the results are questionable.
- Most procedures utilize less information from the sample data (e.g. distribution) and therefore the test are in general less powerful.
- There is no general underlying principles and therefore the procedures are not widely applicable. Nonparametric statistics are designed for specific problems and data sets.
- The procedures do not solve the problem of robustness that classical least square method suffers from but rather they are different methods for different purposes. Unlike robust estimators, most of the techniques in nonparametric analyses are for hypothesis testing but not for the purpose of location estimation. Therefore, nonparametric statistics are not substitute for classical parametric approach.

Advantages of Modern Robust Statistics:

- The methods are developed with the general principle of achieving robustness without sacrificing power. This modern robust statistics seeks to provide methods as substitutes for or alternatives to classical methods based on the dogma of normality. Therefore, modern robust statistics is expected to be more widely applicable than nonparametric statistics.
- Because robust procedures take into account the distribution of the data (whatever the shape), robust procedures in general produce more powerful tests than nonparametric procedures. While robust procedures undertake to deal with non normality and outlying value problems, nonparametric analyses are designed to avoid the problems.
- Modern robust procedures are also concerned with parameter or location

estimation while nonparametric statistics is not. Therefore, modern robust statistics has wider applications. In this regard, the best that can be said is that nonparametric analyses are a subset of a more general family of robust procedures.

- Like nonparametric statistics, because of their robustness, the hazards of inappropriate application are less consequential than the classical parametric procedures.

Disadvantages of Modern Robust Statistics:

- In terms of practical application, nonparametric analyses seem to favoured because of their popularity and ease of use. Functional fixation with normal model and sheer resistance to change might hinder the wide application of the robust techniques.
- The procedures generally involve iteration and complex formulas so they need more computing time and thus are more costly than nonparametric techniques.
- As Mallows (1979) observes, robust procedures are somewhat novel and unfamiliar to many clients and thus can pose an obstacle to effective use of the methods.

WHAT METHOD TO USE

What method to use on a particular data analysis depends on the purpose of the study and the availability of tools. The variety of situation sometimes dictate particular methods.

- If the purpose of the study is to test characteristics of population and no parameter estimation is involved, nonparametric analyses may be the choice especially if the size of data is small.
- If the assumption of normality for a parametric procedure is not met and some population parameters should be estimated, modern robust methods are the only choice.
- If the data are count or rank measurements, nonparametric procedure would be the choice because of ease of use or the only procedures available.
- When location estimation is involved and the sample size is large, modern robust methods would be the choice because of the power of tests.
- When power of test is more important than practicality, modern robust method is better than nonparametric procedures.

The more difficult decision is to choose between classical and robust procedures. In general, it is not a good idea to blindly apply a model prior to sampling and then make statistical inference about the distribution characteristics from the sample without worrying whether or not the model is appropriate to achieve the purpose. The characteristics of the sample may have to be investigated and then a broader model that is robust for all possible distributions under consideration can be applied. However, in complicated and large data sets, identifying normality and outliers are often difficult. when this is the case, Hogg (1979) recommends the following steps:

- Perform both the usual least squares analyses and robust procedures.
- If estimates from both methods are in agreement, report the agreement and the usual statistical summaries associated with least square method.
- It the results from both methods are not quite in agreement, investigate the data carefully and pay special attention at the points with large residuals from the robust fit.
- If points with large residuals are suspected, find possible reasons for that (recording error or the points are trying to tell something).

REFERENCES

Andrew, D.F. "A Robust Method for Multiple Linear Regression." *Technometrics*, November 1974, v. 16, pp. 523-531.

Booth, David E. "Robust Regression and Problem Banks." *UMAP Modules Tools for Teaching*. Comap, Inc., 1986. pp. 183-215

Daniel, Wayne W. *Applied Nonparametric Statistics*. Boston: PWS-Kent Publishing Co,. 1990.

Dielman, Terry E. *Applied Regression Analysis for Business and Economics*. Boston: PWS-Kent Publishing Company, 1991.

Elliott, John, *et al*. "The Impact of SFAS No. 2 on Firm Expenditures on Research and Development: Replications and Extensions." *Journal of Accounting Research* (Spring 1984), pp. 85-102.

Financial Accounting Standards Board (FASB). *Accounting for Research and Development Costs* (FASB, 1974).

Francis, Jere R. and Sara Ann Reiter. "Determinants of Corporate Pension Funding Strategy." *Journal of Accounting and Economics* 9 (1987), pp. 35-59.

Grabowski, Henry G. "The Determinants of Industrial Research and Development: A Study of the Chemical, Drug and Petroleum Industries." *Journal of Political Economy* (March/April 1968), pp. 293-305.

Hogg, Robert V. "Statistical Robustness: One View of Its Use in Application Today." The American Statistician, August 1979, v. 33, pp. 108-115.

Horwitz, Bertrand and Richard Kolodny. "The FASB, the SEC and R&D." *The Bell Journal of Economics* (Spring 1981), pp. 249-262.

Huber, Peter J. "Robust Statistics: A Review." The Annal of Mathematical Statistics, 1972, v. 43, No. 4, pp. 1041-1067.

Madden, Donald L., Levis D. McCullers, and Relmond P. VanDaniker. "The Materiality of Research and Development Expenditures." *Journal of Accounting Research* (Autumn 1972), pp. 417-420.

Mallowa, Colin L. "Robust Methods-Some Examples of Their Use." The American Statistician, November 1979, v. 33, pp. 179-184.

Malmquist, David H. "Efficient Contracting and the Choice of Accounting Method in the Oil and Gas Industry." *Journal of Accounting and Economics* 12 (1990), pp. 173-205.

Rey, William J. "Robust Statistical Methods." Lecture Notes in Mathematics. New York: Springer-Verlag, 1978.

Rousseeuw, P. "Least Median of Squares Regression." *Journal of the American Statistical Association* 79 (1984), pp. 871-880.

Shehata, Mohamed. "Self-Selection Bias and the Economic Consequences of Accounting Regulation: An Application of Two-Stage Switching Regression to SFAS No. 2." *The Accounting Review* (October 1991), pp. 768-787.

Staudte, Jr., Robert G, "Robust Estimation." Queen's Papers in Pure and Applied Mathematics. Kingston: Queen's University, Ontario, 1980.

- Suwardjono. "Self-Selection Bias Model: An Application of Two-Stage Switching Regression to SFAS No. 19." *Journal of Indonesian Economy & Business*, Vol. 16 No. 3 (July, 2001).